Environmental Education Research
Vol. 17, No. 1, February 2011, 91–111

Routledge
Taylor & Francis Group

Development and validation of scales to measure environmental responsibility, character development, and attitudes toward school

Robert Baxter Powell[a,b]*, Marc J. Stern[c], Brian D. Krohn[d] and Nicole Ardoin[e]

[a]*Department of Parks, Recreation, and Tourism Management, Clemson University, 281 Lehotsky Hall, Clemson 29634, USA;* [b]*Department of Forestry and Natural Resources, Clemson University, 281 Lehotsky Hall, Clemson 29634, USA;* [c]*Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, USA;* [d]*Department of Tourism, Conventions, and Event Management, Indiana University-Purdue University-Indianapolis, Indianapolis, USA;* [e]*School of Education & Woods Institute for the Environment, Stanford University, 485 Lasuen Mall, Stanford 94305, USA*

(*Received 14 August 2009; final version received 8 February 2010*)

This investigation examines the use of structural equation modeling (SEM) procedures to develop and validate scales to measure environmental responsibility, character development and leadership, and attitudes toward school for environmental education programs servicing middle school children. The scales represent outcomes commonly of interest to environmental education programs and also to after-school and positive youth development activities. First, we developed the scales using confirmatory factor analysis (CFA) and then we used multi-group longitudinal CFA to cross-validate the model with data collected before participation in the environmental education program, immediately after the program, and three months later. The results support a three-factor model, producing three scales that appear to be valid and reliable.

Keywords: scale development; structural equation modeling; measurement invariance; environmental responsibility; character development; attitudes toward school; environmental education

Introduction

Although environmental education programs vary in their specific objectives, certain outcomes crosscut the field. Some of the most typical outcomes are associated with enhancing participants' environmental knowledge, attitudes, and behaviors (Hungerford and Volk 1990; Leeming, Dwyer, and Bracken 1995; Musser and Malkus 1994; Stern, Powell, and Ardoin 2008). In addition, many programs and experts acknowledge the great potential and importance of environmental education programs for influencing participants' sense of empowerment, leadership skills, overall attitudes toward school, and elements of social skills and character development (Arvai et al. 2004; Athman and Monroe 2004; Carr 2004; Ernst and Monroe 2006; Lieberman and Hoody 1998; Linney 2007; Morgan et al. 2009; Simmons et al. 2004).

In this paper, we examine the development and validation of three scales to measure environmental responsibility, character development and leadership, and

*Corresponding author. Email: rbp@clemson.edu

ISSN 1350-4622 print/ISSN 1469-5871 online

DOI: 10.1080/13504621003692891
http://www.informaworld.com

attitudes toward school for environmental education programs serving middle school children. We address issues of theoretical and methodological importance by examining the conceptual foundation, development, construct validity and reliability, and psychometric qualities of these scales through the use of confirmatory factor analysis (CFA), which is a form of structural equation modeling (SEM). SEM is a group of rigorous statistical procedures that uses a confirmatory (as opposed to exploratory) approach, provides clear statistical indices for making decisions, and explicitly models the relationships between items, the concepts they represent, and the unexplained residual variance or measurement error (Kline 2005). First, we developed the scales using CFA and then we used multi-group longitudinal CFA to cross-validate the model. To this end, we performed SEM on data collected before participation in the environmental education program, immediately after the program, and three months later.

Study site

The NorthBay Adventure Center (NorthBay) is located on the shores of the Chesapeake Bay in northeastern Maryland. During the school year, NorthBay provides five-day residential programs with an audience consisting primarily of middle school groups. Each year NorthBay services more than 7000 students from Maryland public and private schools, with groups ranging in size from 50 to 350 students. Each group is accompanied by teachers and adult chaperones. Although NorthBay also offers a follow-up after-school program, the data reported herein only relate to on-site programming.

NorthBay's on-site programming focuses on three core themes: environmental education, adventure, and character development. NorthBay's environmental education curriculum has been developed in conjunction with and is endorsed by the Maryland State Department of Education and is based on Hungerford et al.'s (2003) 'investigating and evaluating environmental issues and actions' (IEEIA) model. The IEEIA model provides students with a hands-on opportunity to investigate environmental issues using a critical scientific approach and addressing multiple perspectives pertaining to environmental problems. NorthBay's environmental education program uses the five ecosystems available on and around the campus as an outdoor classroom. The adventure component of the NorthBay experience is designed to complement the education program and seeks to challenge and empower students with the goal of enhancing self-confidence, communication, team work, and leadership. This goal differs from the program goal associated with another component of this program, the use of IEEIA. The character development programming is embedded into all aspects of the NorthBay experience but is highlighted during evening programs. These programs use multi-media and live performance to address what NorthBay refers to as 'future killers,' which include dropping out of school and using drugs, among others. In addition to focusing on the students' abilities to make their own informed decisions about their futures, the evening program also reinforces environmental themes introduced during the day by drawing analogies between environmental lessons and challenges at home. For example, the program links the idea of wetlands as ecological filters for pollutants in nature and role models as filters for negative influences in students' lives. Similar analogies are drawn regarding ecological concepts such as invasive species and (bio)diversity. Overall, the program content reflects NorthBay's mission statement:

> Challenge middle school students to realize that their attitudes and actions have a lasting impact on their future, the environment and the people around them by using approved Maryland Department of Education curriculum and the outdoors as an integrating context. (http://www.northbayadventure.com/)

Conceptual framework and scale development process

Developing the NorthBay evaluation system was a participatory process, engaging a range of staff members – including managers, educators, and operations personnel – in a year-long planning process. This planning process largely followed the Sustainable Evaluation Program framework development process (Powell, Stern, and Ardoin 2006), which incorporates elements of utilization-focused (Patton 1996), participatory (Cousins and Earl 1992; King 1998), theory-driven (Campbell and Stanley 1963; Rossi and Freeman 1993; Weiss 1998), and consumer-based (Bledsoe and Graham 2005) approaches. In this process, the programmatic goals and their operational definitions provided the conceptual framework and drove the subsequent development of specific items for measuring each of the desired outcomes. During the planning process, NorthBay administrators, staff, and stakeholders identified three outcomes supported by five programmatic goals pertaining to character development and leadership, environmental stewardship, and attitudes toward school (Table 1).

Using these three outcomes and their corresponding goals and specific objectives, we developed and refined the survey items to best measure each concept following procedures outlined by DeVellis (2003). Because each outcome of interest had a specific operational definition with multiple objectives, we constructed scales using multiple question types to best reflect each construct. We also varied question types to alleviate testing fatigue, as children have been found to have shorter attention spans than adults (e.g., Nelson et al. 1999; Ruff and Lawson 1990). Although there are benefits to using multiple question types within a scale, method and measurement error may be introduced (Byrne 2006). This potential for the introduction of method and measurement error will be explicitly addressed in the Results section of this paper. The three primary outcomes of interest are described in further detail below.

Character development and leadership

NorthBay staff identified two primary goals relating to character development and leadership. Three main objectives were associated with each goal (Table 1). As conceptualized by the NorthBay staff, character development and leadership focus on participants' feelings of self-empowerment, confidence, and competence and their relationships with others. Although the environmental education literature considers these to be important general themes (e.g., Carr 2004; Morgan et al. 2009; Simmons et al. 2004), NorthBay's goals in this respect are more clearly articulated in the after-school programming literature (NIOST 2008). In particular, the concepts are closely related to those commonly referred to as positive youth development (PYD). PYD outcomes typically include the development of social and decision-making skills, caring adult and peer relationships, self-confidence, positive self-images or identities, a sense of caring and compassion for others, and a sense of right and wrong (Klein et al. 2006; Lerner et al. 2005; Smith 2007). Similarly, fields such as outdoor and experiential education also share similar outcomes, though we were unable to locate scales that were directly and specifically useful for this context. Therefore, we

Table 1. Outcomes, goals, and objectives of the NorthBay program.

Character development and leadership

Goal 1. To empower NorthBay participants to recognize they can make choices that can lead to a promising future

Primary objectives:

- Students believe that their personal choices have an impact on their future.
- Students believe they can make appropriate, positive life choices.
- Students develop and rely on a support structure.

Goal 2. To foster the growth of leadership skills in NorthBay participants

Primary objectives:

- Students care about others.
- Students clearly communicate with one another.
- Students take responsibility for their actions.

Environmental responsibility

Goal. To inspire environmental responsibility and community respect in NorthBay adventurers/explorers

Primary objectives:

- Students understand that their actions affect the environment.
- Students will be involved in making a difference for the community.
- Students will engage in environmental stewardship.
- Students recognize environmental health, good and bad.

Attitudes toward school

Goal 1. To inspire improved academic performance of NorthBay participants

Primary objectives:

- Students improve academic performance.
- Students are more excited about learning.
- Students are more involved in extra-curricular activities.

Goal 2. To encourage a healthy and empowered culture in participant schools

Primary objectives:

- Students are enthusiastic, energetic learners who are respectful of others, the school, and the community.
- Students perceive teachers as supportive.
- Students expect themselves to complete school/education.

developed nine survey items to construct the scale (see Table 2). All questions were measured on five-point scales. Four questions used answer choices of 'strongly agree, agree, neutral, disagree, and strongly disagree,' and five questions used answer choices of 'always, often, sometimes, hardly ever, and never.'

Environmental responsibility

NorthBay staff incorporated traditional themes of environmental awareness and stewardship in their definition of environmental responsibility. They also included students' communities as part of the environment and their concern for these communities in their definition (Table 1). Six survey items were developed to construct the environmental responsibility scale (Table 2). All questions were measured on five-point scales. Three questions used answer choices of 'strongly agree, agree, neutral, disagree, and strongly disagree'; two questions used answer choices of 'very interested, pretty interested, a little interested, hardly interested, and not interested'; and

one question used answer choices of 'always, often, sometimes, hardly ever, and never.' Other studies examining similar environmental responsibility outcomes included: Bunting and Cousins' (1983) Children's Environmental Response Inventory; Evans et al.'s (2007) study of children's environmental attitudes and behaviors; Leeming, Dwyer, and Bracken's (1995) children's environmental attitude and knowledge scale; Musser and Malkus' (1994) children's attitudes toward the environment scale; and Stern, Powell, and Ardoin's (2008) children's environmental stewardship scale. However, as reported in the literature, none of those studies used SEM techniques to validate these scales.

Attitudes toward school

The 'attitudes toward school' construct reflects NorthBay's desire to positively influence students' academic motivations and their perceptions of their school (Table 1). This conceptualization incorporates the ideas of other authors who have worked to gauge the impact of environmental education initiatives on attitudes toward school. For example, Lieberman and Hoody (1998) showed that environmentally-based curricula can promote student gains in academic motivation and standardized test scores and reduce classroom management problems. Ernst and Monroe studied the impacts of 12 environment-based education programs in Florida on students' critical thinking skills, dispositions toward critical thinking, and academic motivation, finding positive connections with each (Athman and Monroe 2004; Ernst and Monroe 2006). Similarly, Hungerford, Volk, and Ramsey (2000) explored the value of IEEIA programs on citizenship and academic achievement and also found positive correlations. The after-school literature has further examined the connections between out-of-school programs and attitudes toward school, examining outcomes such as student grades and performance on standardized tests, behavior in the classroom, emotional attachment to the school, motivation and engagement in learning, and homework effort and completion (Durlak and Weissberg 2007; Miller 2005). Positive outcomes have been linked to positive and respectful relationships between teachers and students and among peers, involvement in extra-curricular activities, and comfortable learning environments, among other variables (Klein et al. 2006; Miller 2005). Thus, NorthBay's conceptualization of attitudes toward school aligns with current thinking about what generates positive student outcomes.

Ten survey items were developed to construct the scale (Table 2). All questions were measured on five-point scales: three questions used answer choices of 'strongly agree, agree, neutral, disagree, and strongly disagree'; three questions used answer choices of 'always, often, sometimes, hardly ever, and never'; two questions used answer choices of 'very interested, pretty interested, a little interested, hardly interested, and not interested'; and two proportion questions used answer choices of 'all of them, more than half, about half, less than half, none of them.'

Methods

Sample and data collection

The evaluation-development process was completed in August 2006. The survey was pilot tested during the 2006–2007 academic year. At that time, we used exploratory factor analysis to further refine the operationalization of key concepts, following

Table 2. Means, standard deviations, question type, and hypothesized factors of observed variables.

Conceptual framework and items	Code	Question type	Before *M*	Before SD	After *M*	After SD	Follow-up *M*	Follow-up SD
Character development and leadership								
The choices I make today can change my entire life.	Choice	Level of agreement	3.58	1.05	4.01	1.03	3.99	1.00
I have people who support me when I need help.	Support	Level of agreement	4.37	0.87	4.38	0.84	4.42	0.88
I (will) make up excuses when I behave badly.	Excuse	How often?	2.68	1.08	2.62	1.10	2.86	1.06
I (will) take responsibility for my mistakes.	Mistake	How often?	3.87	1.06	4.05	1.00	3.75	1.03
I (will) talk to my friends about making positive life choices.	Life change	How often?	3.22	1.30	3.70	1.22	3.41	1.28
I feel comfortable arguing for what I believe in.	Argue	Level of agreement	3.66	1.20	3.75	1.16	3.73	1.12
I can be a good leader.	Leader	Level of agreement	4.13	0.99	4.11	0.99	4.04	1.06
I (will) help my friends with their problems.	Help	How often?	3.95	1.05	4.15	0.96	4.08	1.00
I (will) talk to my family or friends outside of school about what I've learned.	Talk	How often?	3.16	1.23	3.33	1.13	3.20	1.14
Environmental responsibility								
My actions impact the health of the environment.	Environment	Level of agreement	3.44	1.09	3.78	1.01	3.63	1.00
I have the power to help protect the environment.	Protect	Level of agreement	3.84	1.08	3.98	1.04	3.89	1.05
I can make a change in my community.	Change	Level of agreement	3.83	1.04	3.84	1.02	3.79	1.05
Learning about how to protect the environment.	Learning environment	Level of interest	3.77	1.13	3.76	1.15	3.63	1.15
Working to make my community a better place.	Better	Level of interest	4.00	1.06	3.94	1.08	3.81	1.13
I (will) work as a volunteer in my community.	Service	How often?	2.43	1.29	3.11	1.23	2.64	1.24
Attitudes toward school								
Going to school is a waste of time for me.	Waste	Level of agreement	1.64	0.94	1.64	0.94	1.71	1.03
I enjoy school.	School	Level of agreement	3.57	1.14	3.58	1.14	3.47	1.23
Learning about new subjects in school.	Subject	Level of interest	3.76	1.04	3.67	1.11	3.68	1.10

Table 2. (*Continued*).

Conceptual framework and items	Code	Question type	Before		After		Follow-up	
			M	SD	*M*	SD	*M*	SD
Going to college.	College	Level of interest	4.55	0.89	4.58	0.84	4.60	0.82
I (will) research things that I am curious about.	Research	How often?	3.38	1.24	3.51	1.19	3.37	1.23
I (will) complete all of my schoolwork on time.	Work	How often?	4.23	0.88	4.44	0.84	4.07	0.96
My school is a dangerous place.	Danger	Level of agreement	1.67	0.97	1.92	1.07	2.15	1.14
How many of your teachers really care about you?	Teacher care	Proportion	4.38	0.95	4.30	1.02	4.10	1.16
How many of your teachers believe that you can succeed?	Teacher success	Proportion	4.65	0.73	4.54	0.85	4.41	1.00
I (will) pay attention to the teacher in class.	Attention	How often?	4.19	0.95	4.31	0.94	3.98	0.99

DeVellis (2003) and Presser et al. (2004). During the 2007–2008 school year, NorthBay staff administered surveys to all students participating in NorthBay programs over three specific weeks (*n* = 868). Students completed surveys immediately on arrival to the NorthBay campus and immediately before departure, during their final meal on the NorthBay campus. Response rates were at or near 100% for on-campus surveys. Research occurred during these three weeks based on the attendance of three middle schools – one urban public school, one rural public school, and one suburban/rural public school. Using a purposive sampling method, the researchers selected the students of these schools because they represented a cross-section of the rural-to-urban spectrum as well as ethnic and socioeconomic composition. To increase the sample size and reduce sampling bias, all students visiting NorthBay during the same three weeks as the three selected schools were administered surveys. Students from 18 middle schools were included in the 2007–2008 study. The other 15 middle schools in the 2007–2008 research included 14 primarily urban public schools and one rural public school.

Three months after their program, follow-up surveys were administered to students at seven of the schools. NorthBay staff were not successful in arranging follow-up surveys with the other schools. The surveys were administered in the classroom by teachers, who then returned the completed surveys to NorthBay. This yielded 349 follow-up respondents. Not all students who participated in the before and after surveys were present in class on the day of the follow-up survey.

Structural equation modeling

SEM is a group of advanced statistical procedures that are becoming increasingly popular for multivariate analysis, and was employed to analyze and validate the scales used in this study. It is being used more frequently because it is confirmatory (as opposed to exploratory) in nature and requires the researcher to have an explicit model; it explicitly models observed as well as latent variables (the actual items and the concepts they are measuring); it can model and correct measurement error which reduces inaccuracies; and finally, it allows for the analysis of a complete multivariate model including direct and indirect effects (Byrne 2006; Kline 2005). In SEM, measurement error refers to 'indicator variance not explained by the factors' (Kline 2005, 73).

We used the EQS v6.1 software (Bentler 2005) to perform the statistical analyses, which progressed in several stages. First, the data were screened for univariate and multivariate deviations from normality. We also conducted tests looking for patterns in the missing data. Next, the structure of the items was tested using the follow-up data in CFA. This is an explicit test of the hypothesized model regarding which items would relate to each latent construct (character development and leadership, environmental responsibility, and attitudes toward school) and is an iterative process in which diagnostics indicate potential modifications to the model until the best fit is identified. Lastly, the final model was cross-tested for validity by using tests of measurement invariance through SEM. This analysis determines if the hypothesized model, including the items comprising a scale and the hypothesized factor structure of the scales, is stable across different samples (e.g., Kyle, Graefe, and Manning 2005; Whiteside-Mansell and Corwyn 2003). By using longitudinal data collected pre-program, post-program, and three months post-program to cross-validate our model, we provide a rigorous analysis that confirms the stability of the scales and their psychometric qualities (Byrne, Shavelson, and Muthen 1989).

Data preparation

Initially data were screened independently using SPSS V.13 for cases missing more than 50% of data and univariate outliers (Tabachnick and Fidell 2007). No cases exceeded recommended cut-off values (Fox 1991). The primary analysis technique used in this paper to analyze the structure of the items is CFA using SEM with maximum likelihood. As Yuan and Bentler (2001) and Hatcher (1994) describe, this technique is sensitive to variation in the normality of the data. Particularly, non-normal multivariate kurtosis (variance due to infrequent extreme deviations) can lead to improper results or parameter estimation. Therefore, following the procedures outlined by Hatcher (1994), the data collected for each of the time periods were examined for influential cases. The EQS software package (EQS 6.1) provides measures for cases that have the most contribution to multivariate kurtosis. After analysis, a number of cases that exceeded the threshold were removed from each of the three datasets: 14 cases in the pre-program data, 13 cases in the post-program data, and 2 cases in the follow-up data. The removal of these cases improved the overall multivariate normality, thus reducing the possibility of erroneous results. Where appropriate, robust measures, which correct for non-normality in the data, were also used to ensure correct estimation of the parameters (Bentler and Yuan 1999; Byrne 2006).

Tests of missing data

We identified missing data and examined those missing data points for patterns of ‘missingness’ (Schafer and Graham 2002). We undertook a test of ‘missing completely at random’ (MCAR) (Little and Rubin 1987) to examine the missing data to see if those data were related to responses on other items. This procedure was conducted independently for each dataset. There was significant evidence in all three groups that a pattern existed in the missingness of the data (Allison 2003). Because of the significance of the missing patterns, we did not impute the missing data. Imputation refers to a group of statistical procedures that analyzes an individual respondent’s scores and ‘imputes’ or infers replacement values for missing data. Rather, cases with missing data were excluded from the analysis. The number of cases used in the following analyses were as follows: pre-program data = 765; post-program data = 760; and follow-up data = 322 (Table 3).

Determining factor structure using confirmatory factor analysis

We used confirmatory factor analysis (CFA) to investigate and refine the NorthBay scales because the analysis explicitly tests a hypothesized measurement model (as opposed to an exploratory approach), accounts for sources of common measurement and method error that is inherent in survey research, and provides empirical justification

Table 3. Sample sizes and missing cases.

Sample	Sample size	Missing cases	Numbers used for analysis
Pre-experience	868	103	765
Post-experience	868	108	760
Follow-up (three months after)	349	27	322

for model and scale development decisions (Byrne 2006; Kline 2005; Noar 2003). CFA also provides multiple statistics that can be used to evaluate the appropriateness (goodness-of-fit) of a specified model to the sample data (Byrne 2006). In this paper, we report Satorra-Bentler Scaled Chi-Square (S-Bχ^2), Robust Comparative Fit Index (CFI), Standardized Root Mean Square Residual (SRMR), Robust Root Mean Square Error of Approximation (RMSEA), and its associated 90% confidence interval. Where appropriate, the non-robust chi-square value (χ^2) is also presented for comparison/calculation of variation between models (Byrne 2006; Kline 2005).

The S-Bχ^2, which should be interpreted like a χ^2, is reported because it corrects for the degree of kurtosis in the data (Satorra and Bentler 1994). The CFI is an 'incremental or comparative fit index' that evaluates the change in fit between the hypothesized model and the 'independence model' (Byrne 2006, 97; Kline 2005, 140). The independence model assumes that all the variables in the model are unrelated. The CFI represents the total covariation in the data and is measured on a scale of 0 to 1 with values greater than .9 indicating an acceptable fit and values greater than .95 indicating an excellent fit (Byrne 2006; Hu and Bentler 1999). Following the advice of Byrne (2006) and Bentler (1990), we report the Robust CFI which accounts for non-normality of data. The SRMR statistic provides the average difference between the sample and the predicted correlation matrices and, thus, is not susceptible to non-normality (Byrne 2006). The SRMR uses standardized values with the range of scores between 0 and 1; values less than .1 are considered acceptable and less than .05 are considered a good fit (Hu and Bentler 1995; Kline 2005). The RMSEA is based on the average lack of fit per degree of freedom; therefore, as the fit improves, the RMSEA decreases. As such, this measure is sensitive to the degrees of freedom and the complexity of the model (Byrne 2006). Like the SRMR, the scores range between 0 and 1, with values of .05 to .08 deemed acceptable and values less than .05 considered excellent (Browne and Cudeck 1993; Hu and Bentler 1999). In this analysis, we report the Robust RMSEA to account for non-normality in the data. To explore potential adjustments to the measurement model, the Lagrange Multiplier (LM) test was used (Byrne 2006; Kline 2005). This test identifies items and inter-item covariance that degrade the model and that, if removed, would improve the overall fit. In other words, this test suggests the most statistically valid adjustments to the model that will improve its goodness-of-fit. Because the LM test is empirical, model adjustment decisions also included theoretical justification (Byrne 2006; Tabachnick and Fidell 2007).

Results

Of the 868 students surveyed immediately before and after the NorthBay program, 55% were female, and students' ages ranged from 10 to 15, with a mean and median age of 11 years. Of the 350 students surveyed three months after the NorthBay program, 54% were female, and students' ages ranged from 10 to 15, with a mean and median age of 11 years.

Consistent with recommended CFA covariance structure analysis procedures, we explored our conceptually based hypothesized factor structure as well as alternative models (Breckler 1990; Browne and Cudeck 1993). Our first step in determining the structure was to test the items as a single-factor structure (Model 1). This model assumes that all survey items contribute to a single latent construct. The resulting fit was poor (S-Bχ^2 = 843.71; CFI = .761; SRMR = .071; RMSEA = .081) indicating that the model does not fit well into a one-factor model.

As described previously, the survey instrument included four different types of questions (‘strongly agree’ to ‘strongly disagree’, level of interest, proportion, and behavioral intention questions with answer choices ‘never’ to ‘always’), and each hypothesized scale contained multiple types. Scales are often constructed to measure a factor such as ‘environmental responsibility’ from items using the same Likert-type response choices stemming from the same initial survey question. In those cases, no explicit testing for method/measurement bias is necessary with CFA because the potential error is modeled and accounted for within the factor structure. The different question formats used in the NorthBay survey, however, could lead to error in the model, as variation in results may be attributed to the respondents’ reaction to the question type rather than the latent construct that the question is attempting to measure. In Model 2, we modeled this potential bias into the structure. The model’s fit increased (S-Bχ^2 = 590.29; CFI = .860; SRMR = .058; RMSEA = .065), which signifies that the question type had an influence on students’ responses. Despite the improvement, this model still generated poor statistical fit.

For the next model (Model 3), we used the five goals of the project, which supported the three overarching outcomes to derive a five-factor structure. The fit statistics improved over the one-factor model, but the results indicated that the model structure could still be improved (S-Bχ^2 = 450.08; CFI = .914; SRMR = .109; RMSEA = .052). Finally, we constructed a three-factor model, which reflects the hypothesized outcomes of the NorthBay program and the intended structure of the items. The results of the three-factor model (Model 4) revealed an improvement in fit (S-Bχ^2 = 361.64; CFI = .952; SRMR = .047; RMSEA = .039) signifying that the three-factor model was more appropriate than each prior model.

Model 4, the three-factor model, served as the base model and starting point for refining the measurement model using the procedures outlined by Hatcher (1994). First, we examined the factor loadings to uncover insignificant paths/low loadings and error terms for high residual/correlated errors. EQS reports results from the Wald test which identifies items within the model that if dropped, would increase the overall fit. The software also reports the results from the LM test, which identifies additional covariance relationships not reported in the model. Hatcher (1994) suggests that a satisfactory model should: (1) display acceptable fit in all relevant indices, (2) contain only significant factor loadings, and (3) contain only items with an acceptable level of normalized residual.

To refine the model and improve the level of fit, the items ‘argue’ and ‘excuse’ were dropped from the leadership/character development factor because of low factor loadings. Similarly, ‘danger’ was dropped from the attitudes toward school factor. These low factor loadings indicate that the items did not measure the target factor as intended. Two additional items, ‘work’ and ‘help,’ were dropped because of higher loadings with the method bias factor than with the specified factor. This indicates that the responses to these items were more highly associated with the question type than with the factor. One item, ‘research’, was hypothesized as measuring academic performance, but demonstrated cross-loading with another factor in the model, ‘character development/leadership’. This item exhibited a stronger empirical link to the ‘character development and leadership’ factor. After review, we felt the item also conceptually aligned with other items in the ‘character development and leadership’ factor because of the item’s emphasis on personal initiative and self-empowerment. We therefore changed its position in the model. Results from Model 5 indicate a substantial improvement in fit (S-Bχ^2 = 183.68; CFI = .981; SRMR = .038; RMSEA = .029).

The Wald and LM tests did not identify additional opportunities for improvement in this model; thus Model 5 represents the final measurement model (Figure 1). Table 4 displays a comparison of fit statistics for each of the models. Finally, internal consistency, also known as reliability testing, was measured for each scale using the Rho coefficient. The Rho coefficient is more appropriate than the Cronbach's alpha in this context because it adjusts for multiple factors, unequal factor loadings, and the use of error terms in CFA (Byrne 2006). The scores ranged from .662–.795 (Table 7). The Rho coefficient should be interpreted in the same way as the Cronbach's alpha with scores above .6 deemed adequate for group prediction (Gay 1991).

Testing validity of the structure

To further explore the validity and psychometric properties of the final model, we used multi-group tests of measurement invariance to examine the stability of the structure across the three data collection points (pre-program, post-program [immediately after program], and follow-up [three months after program]). These cross-validation procedures provide a rigorous analysis that determines whether the items comprising a scale, the hypothesized factor structure of the scales, and their measurements are stable across different samples (Byrne 2006; Byrne, Shavelson, and Muthen 1989). We developed the final measurement model (Model 5) using the follow-up data, which served as Group 1. The post-program data served as Group 2, and the pre-program data served as Group 3. Following recommendations by Vandenberg and Lance (2000), we used increasingly stringent analyses to investigate invariance across multiple samples.

First, we compared the fit statistics from each dataset independently using the identical model (the same configuration). The results indicate that the statistics are roughly the same with most of the differences attributed to the differences in sample size (Table 5). Second, we developed a multi-group configural model that simultaneously tested the structure across the three samples. The configural test of invariance simultaneously compares the 'number of factors and factor-loading pattern' across the groups (Byrne 2006, 233). Third, we tested the measurement invariance by constraining the free factor loadings (loadings that are freely estimated and not fixed to 1 for identification and scaling) to be equal. In this step, the pattern between factor loadings of items and factors in the base model are compared against the two additional samples by constraining them to be equal (Byrne 2006). Lastly, we conducted the structural invariance test, which compares the relationships between factor and error covariances by constraining them to be equal across the three samples (Byrne 2006).

For each step, we tested the significance of the difference between each of the models using a procedure that adjusts the chi-square difference test to the Satorra-Bentler robust measures. When interpreting the results of this test, a *p*-value of less than .05 indicates that the covariance structures of the models differ significantly and, therefore, were not stable across the samples (Byrne 2006). Table 6 lists the fit statistics of each analysis as well as the adjusted chi-square difference tests. The results of the multi-group configural test produced statistics indicative of a good fitting model and suggests that the structure of factors and items (the factor structure and factor-loading patterns) are the same across the three samples (S-Bχ^2 = 689.43; CFI = .970; SRMR = .036; RMSEA = .031). When the factor loadings were constrained to be equal, the results indicated a marginally significant difference between the measurement of the

Table 4. Results from model building and adjustments.

Model development	df	S-Bχ^2	CFI	SRMR	RMSEA[a]	Δdf	ΔS-Bχ^2	*p*	CFI change	Δdf
Model 1: single-factor model	275	843.71	0.761	0.071	0.081 (.075, .087)	—	—	—	—	—
Model 2: single-factor model with measurement error removed	252	590.29	0.86	0.058	0.065 (.058, .072)	−23	−253.42	<.01	0.099	−23
Model 3: five-factor model	243	450.08	0.914	0.109	0.052 (.044, .059)	−32	−393.63	0.04	0.153	−32
Model 4: three-factor model	246	361.64	0.952	0.047	0.039 (.030, .047)	−29	−482.07	0.50	0.191	−29
Model 5: adjusted three-factor model	145	183.68	0.981	0.038	0.029 (.013, .041)	−130	−660.03	0.01	0.22	−130

[a]95% confidence interval around the RMSEA.

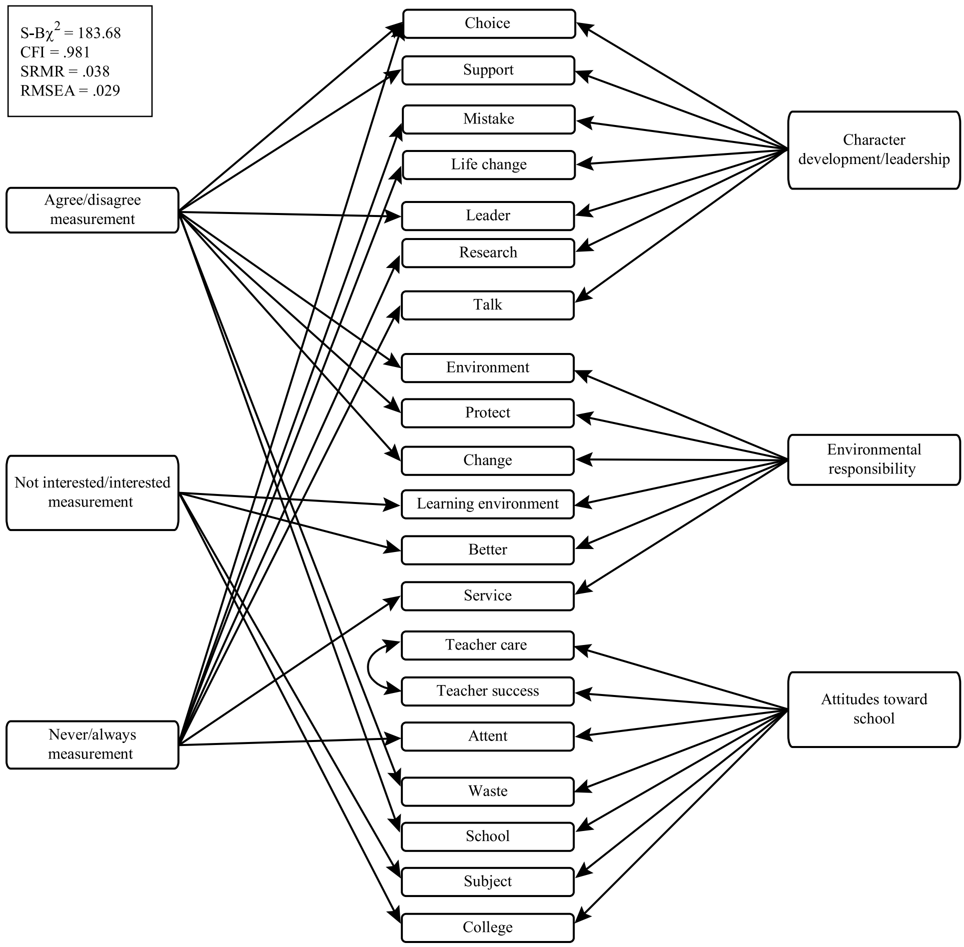

Figure 1. Model 5: adjusted three-factor model.

models (adjusted chi-square difference = 49.40; df = 34; p = .046). When covariances were constrained to be equal, the multi-group test of measurement invariance also showed a significant difference between groups (adjusted chi-square difference = 63.40; df = 46; p = .039). This suggests that while the structure was stable, one or more of the metrics performed differently across time or between groups.

Further post-hoc analysis using the LM test, which indentifies the sources of differences in the comparison tests, indicated that the 'service' item performed differently

Table 5. Comparison of group fit statistics.

Model	df	S-Bχ^2	CFI	SRMR	RMSEA[a]
Follow-up data	145	188.81	0.978	0.039	0.031 (.016, .042)
Post-program data	145	230.63	0.977	0.033	0.033 (.021, .034)
Pre-program data	145	257.52	0.958	0.035	0.035 (.025, .038)

[a]95% confidence interval around the RMSEA.

Table 6. Multi-group tests of invariance and chi-square difference tests of invariance.

Hypothesis tests	df	χ^2	S-Bχ^2	CFI	SRMR	RMSEA[a]	Δdf	ΔS-Bχ^2	p
Multi-group configural model	435	834.23	689.432	0.970	0.036	0.031 (.027, .035)			
Factor loadings constraints	469	896.926	738.632	0.969	0.045	0.031 (.026, .035)	34	49.40	0.046
Covariance constraints	481	914.493	752.442	0.969	0.047	0.030 (.026, .034)	46	63.40	0.039
Factor loading constraints minus service constraint	468	887.297	731.428	0.969	0.044	0.030 (.026, .034)	33	42.33	0.128
Covariance constraints minus service constraint	480	904.321	744.772	0.969	0.045	0.030 (.026, .034)	45	55.86	0.128

[a]95% confidence interval around the RMSEA.

for Group 2 (post) and Groups 1 (follow-up) and 3 (pre). Therefore, we removed the constraints on 'service' (allowing for these loadings to be freely estimated) and then re-analyzed the two measurement invariance tests. This resulted in non-significant differences between the groups for each of the invariance tests (metric invariance adjusted chi-square difference = 42.33; df = 33; $p = .128$; structural adjusted chi-square difference = 55.86; df = 45; $p = .128$). The results from the series of invariance tests suggest that the overall structure of the model is consistent over time. However, the two more stringent metric invariance tests imply that the 'service' item was unstable over time.

Discussion

The purpose of this analysis was to use CFA to investigate the psychometric properties of three scales to measure environmental responsibility, character development and leadership, and attitudes toward school for environmental education programs serving middle school children. We also demonstrated the use of SEM to test measurement invariance over time using three samples. The results indicate that the hypothesized three-factor model is more appropriate than the one-factor or five-factor structures. The results also indicate that the final three scales are valid and reliable (internally consistent) for measuring environmental responsibility, character development and leadership, and attitudes toward school.

We also explicitly addressed method bias by modeling the different types of questions used in the scales. This allowed us to account for and model the error (contribution to variance) caused by different question types and to remove certain items that were highly influenced by this measurement error. Similarly, the series of CFA models provided empirical rationale for removing several items in order to obtain the best possible model. In other words, the final CFA model provides a more parsimonious list of items to measure each of the constructs under consideration: character development and leadership, environmental responsibility, and attitudes toward school. Table 7 provides a list of the final structure and items that constitute the three scales.

After we developed a valid and reliable model, we also conducted increasingly stringent multi-group tests of invariance across the three sample points. These tests of invariance showed that the three scales are valid and reliable over time. The initial tests of invariance demonstrated that the configuration of the model and their fit (number of factors and the factor-loading pattern) are similar across the three groups. However, in the more stringent metric invariance tests, one behavioral intention item from the environmental responsibility scale ('I [will] work as a volunteer in my community') did not perform consistently across the three sampling points. Immediately after the program, the item's factor weighting increased from .373 to .491; in other words, the item more strongly aligned with the concept of environmental responsibility and the other items that measure the concept immediately after participation. Three months later, the factor weighting dropped to .461. A plausible explanation for the item's performance is that the item measured a behavior with a low locus of control and also two different constructs across time. The item, 'working as a volunteer in my community' requires opportunity, which is outside the control of the student; hence, the behavior has a lower locus of control as compared with that of the other behaviors measured in this study such as 'paying attention to the teacher in class,' 'researching things that I am curious about,' or 'talking with my family or

Table 7. Final structure and items constituting scales.

Final structure and items	Factor loading[a]	Question type	Question-type loading[a]
Character development and leadership			
The choices I make today can change my entire life.	.462	Level of agreement	.123
I have people who support me when I need help.	.449	Level of agreement	.060
I (will) take responsibility for my mistakes.	.505	How often?	.328
I (will) talk to my friends about making positive life choices.	.470	How often?	.425
I can be a good leader.	.457	Level of agreement	−.012
I (will) research things that I am curious about.	.502	How often?	.315
I (will) talk to my family or friends outside of school about what I've learned.	.535	How often?	.403
RHO[b] = .751			
Environmental responsibility			
My actions impact the health of the environment.	.434	Level of agreement	.199
I have the power to help protect the environment.	.703	Level of agreement	.214
I can make a change in my community.	.742	Level of agreement	.131
Learning about how to protect the environment.	.700	Level of interest	.351
Working to make my community a better place.	.663	Level of interest	.418
I (will) work as a volunteer in my community.	.442	How often?	.424
RHO[b] = .795			
Attitudes toward school			
Going to school is a waste of time for me.	−.574	Level of agreement	.286
I enjoy school.	.641	Level of agreement	−.409
Learning about new subjects in school.	.644	Level of interest	.448
Going to college.	.457	Level of interest	.220
How many of your teachers really care about you?	.493	Proportion	—
How many of your teachers believe that you can succeed?	.491	Proportion	—
I (will) pay attention to the teacher in class.	.518	How often?	.241
RHO[b] = .662			

[a]Mean standardized loadings across the three samples.
[b]Internal consistency measure that adjusts for unequal factor loadings and error variances (Byrne 2006).

friends outside of school about what I've learned.' Second, the item measures actual behaviors prior to participation, behavioral intentions immediately after the program, and then actual behaviors three months later. This lower level of locus of control coupled with the measurement of intentions post-program may have influenced the performance (the factor weighting and covariance) of this item across time, because many students who intended to work as a volunteer did not perform this behavior three months later. Once the LM test identified this item as not performing the same in the post-program sample as in the follow-up and pre-program groups, and that there was a potential theoretical and practical reason for this, we removed the constraint on this one item (Byrne 2006). The subsequent results indicate that the three scales are reliable and valid across the three data collection points. However, future research should investigate whether locus of control influences the performance of the 'service' item across time.

Conclusion

The SEM procedures described in this paper demonstrate the potential utility of these statistical analyses for assessing the psychometric qualities, validity, and reliability of scales. By using SEM analyses to confirm the three latent constructs and the items that constitute the three scales, and by conducting the subsequent invariance tests to confirm their validity, the three scales may be used in future research as three composite indexes (equally weighted summative scales) (Little et al. 2002; Matsunaga 2008). This will aid environmental education practitioners and researchers interested in similar outcomes who wish to use these scales because composite indexes are easier to create and analyze and the results are easier to interpret (Matsunaga 2008).

The scales developed and validated in this study reflect concepts of importance to a wide range of environmental education and other out-of-school providers. While the environmental responsibility scale represents a new measure of a widely accepted and studied outcome in environmental education, the other two scales address outcomes that have long been considered relevant to environmental education but are infrequently studied in this arena. The 'character development and leadership' and the 'attitudes toward school' scales represent conceptual bridges to outcomes traditionally associated with PYD and after-school programming. The convergence of these different concepts to measure the outcomes of the NorthBay program reflects not only the innovation of the program itself, but also signifies a first step toward acknowledging, quantifying, and evaluating the impact of environmental education programs on additional outcomes of common interest.

Acknowledgments

The authors thank the staff at NorthBay Adventure Center for their commitment to the evaluation program, their visiting students who participated in the surveys, and four anonymous reviewers whose comments improved this paper.

Notes on contributors

Robert Baxter Powell is an assistant professor in the Department of Parks, Recreation, and Tourism Management and the Department of Forestry and Natural Resources at Clemson University. His research focuses on education and communication, protected area management, and sustainable tourism development.

Marc J. Stern is an assistant professor in the Department of Forestry at Virginia Tech. His research focuses on planning processes in environmental organizations and federal agencies, the relations between natural resource management agencies and various publics, and the evaluation of conservation programs.

Brian D. Krohn is an assistant professor in the Department of Tourism, Conventions and Event Management at Indiana University-Purdue University-Indianapolis. His research focuses on decision-making and marketing in sport tourism and the application of statistical methods within the fields of tourism, recreation, and leisure.

Nicole Ardoin is an assistant professor of environmental education at Stanford University. Her research focuses on sense of place and environmental behavior on an ecoregional scale.

References

Allison, P.D. 2003. Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology* 112: 545–57.

Arvai, J.L., V.E.A. Campbell, A. Baird, and L. Rivers. 2004. Teaching students to make better decisions about the environment: Lessons from the decision sciences. *Journal of Environmental Education* 36, no. 1: 33–44.

Athman, J., and M.C. Monroe. 2004. The effects of environment-based education on students' achievement motivation. *Journal of Interpretation Research* 9, no. 1: 9–25.

Bentler, P.M. 1990. Comparative fit indexes in structural models. *Psychological Bulletin* 107: 238–46.

Bentler, P.M. 2005. *Eqs 6 structural equations program manual.* Encino, CA: Multivariate Software. www.mvsoft.com.

Bentler, P.M., and K.H. Yuan. 1999. Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research* 34: 181–97.

Bledsoe, K.L., and J.A. Graham. 2005. The use of multiple evaluation approaches in program evaluation. *American Journal of Evaluation* 26: 302–19.

Breckler, S.J. 1990. Applications of covariance structure modeling in psychology: Cause for concern. *Psychological Bulletin* 107: 260–73.

Browne, M.W., and R. Cudeck. 1993. Alternative ways of assessing model fit. In *Testing structural equation models*, ed. K.A. Bollen and J.S. Long, 445–55. Newbury Park, CA: Sage.

Bunting, T.E., and L.R. Cousins. 1983. Development and application of the children's environmental response inventory. *Journal of Environmental Education* 15: 3–10.

Byrne, B.M. 2006. *Structural equation modeling with EQS: Basic concepts, applications and programming.* 2nd ed. Mahwah, NJ: Erlbaum.

Byrne, B.M., R.J. Shavelson, and B. Muthen. 1989. Testing for equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin* 105: 456–66.

Campbell, D.T., and J.C. Stanley. 1963. *Experimental and quasi-experimental designs for research.* Chicago, IL: Rand McNally.

Carr, D. 2004. Moral values and the arts in environmental education: Towards an ethics of aesthetic appreciation. *Journal of Philosophy of Education* 38, no. 2: 221–39.

Cousins, J.B., and L.M. Earl. 1992. The case for participatory evaluation. *Educational Evaluation and Policy Analysis* 14: 397–400.

DeVellis, R.F. 2003. *Scale development: Theory and applications applied social research methods.* 2nd ed. Thousand Oaks, CA: Sage.

Durlak, J.A., and R.P. Weissberg. 2007. *The impact of after-school programs that promote personal and social skills.* CASEL & W.T. Grant Foundation. http://www.casel.org/downloads/ASP-Full.pdf

Ernst, J., and M. Monroe. 2006. The effects of environment-based education on students: Critical thinking skills and disposition toward critical thinking. *Environmental Education Research* 12, no. 3: 429–43.

Evans, G.W., G. Brauchle, A. Haq, R. Stecker, K. Wong, and E. Shapiro. 2007. Young children's environmental attitudes and behaviors. *Environment and Behavior* 39: 635–58.

Fox, J. 1991. *Regression diagnostics.* Newbury Park, CA: Sage.

Gay, L.R. 1991. *Educational evaluation and measurement: Competencies for analysis and application.* New York: Macmillan.

Hatcher, L. 1994. *A step-by-step approach to using SAS for factor analysis and structural equation modeling.* Cary, NC: SAS Institute.

Hu, L.-T., and P.M. Bentler. 1995. Evaluating model fit. In *Structural equation modeling: Concepts, issues, and applications*, ed. R.H. Hoyle, 76–99. Thousand Oaks, CA: Sage.

Hu, L.-T., and P.M. Bentler. 1999. Cutoff criteria for fit indices in covariance structure analysis: Guidelines, issues, and alternatives. *Structural Equation Modeling* 6: 1–55.

Hungerford, H.R., and T.L. Volk. 1990. Changing learner behavior through environmental education. *Journal of Environmental Education* 21, no. 3: 8–21.

Hungerford, H.R., T.L. Volk, and J. Ramsey. 2000. Instructional impacts of environmental education on citizenship behavior and academic achievement: Research on Investigating and Evaluating Environmental Issues and Actions, 1979–2000. Paper presented at the 29th annual conference of the North American Association for Environmental Education, October 17–21, in South Padre Island, TX.

Hungerford, H.R., T.L. Volk, J.M. Ramsey, R.A. Litherland, and R.B. Peyton. 2003. *Investigating and evaluating environmental issues and actions.* Champaign, IL: Stipes.

King, J.A. 1998. Making sense of participatory evaluation practice. *New Directions for Evaluation* 80: 57–67.

Klein, J.D., P. Sabaratnam, M.M. Auerbach, S.M. Smith, C. Kidjo, K. Lewis, S. Ryan, and C. Dandino. 2006. Development of factor structure of a brief instrument to assess the impact of community programs on positive youth development: The Rochester Evaluation of Asset Development for Youth (READY) tool. *Journal of Adolescent Health* 39: 252–60.

Kline, R.B. 2005. *Principles and practice of structural equation modeling.* 2nd ed. New York: Guilford Press.

Kyle, G., A.R. Graefe, and R.E. Manning. 2005. Testing the dimensionality of place attachment in recreational settings. *Environment and Behavior* 37: 153–77.

Leeming, F.C., W.O. Dwyer, and B.A. Bracken. 1995. Children's environmental attitude and knowledge scale: Construction and validation. *Journal of Environmental Education* 26: 22–33.

Lerner, R.M., J.V. Lerner, J.B. Almerigi, C. Theokas, E. Phelps, S. Gestsdottir, S. Naudeau, et al. 2005. Positive youth development, participation in community youth development programs, and community contributions of fifth-grade adolescents: Findings from the first wave of the 4-H study of positive youth development. *Journal of Early Adolescence* 25, no. 1: 17–71.

Lieberman, G.A., and L.L. Hoody. 1998. *Closing the achievement gap: Using the environment as an integrating context for learning.* San Diego, CA: State Education and Environment Roundtable.

Linney, G. 2007. COEO's Landmark research summary: Reconnecting children through outdoor education. *Pathways: The Ontario Journal of Outdoor Education* 19, no. 2: 33–4.

Little, R.J., and D.B. Rubin. 1987. *Statistical analysis with missing data.* New York: Wiley.

Little, T.D., W.A. Cunningham, G. Shahar, and K.F. Widaman. 2002. To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling* 9: 151–73.

Matsunaga, M. 2008. Item parceling in structural equation modeling: A primer. *Communication Methods and Measures* 2: 260–93.

Miller, B. 2005. *Pathways to success for youth: What count in after-school*. Boston, MA: United Way of Massachusetts. http://supportunitedway.org/files/MARS-Report_0.pdf

Morgan, S.C., S. Hamilton, M.L. Bentley, and S. Myrie. 2009. Environmental education in botanic gardens: Exploring Brooklyn Botanic Garden's Project Green Reach. *Journal of Environmental Education* 40, no. 4: 35–52.

Musser, L.M., and A.J. Malkus. 1994. The children's attitudes toward the environment scale. *Journal of Environmental Education* 25: 22–6.

Nelson, C., L.D. Nugent, E.M. Elliott, I. Ponomarev, and J.S. Saults. 1999. The role of attention in the development of short-term memory: Age differences in the verbal span of apprehension. *Child Development* 70: 1082–97.

NIOST (National Institute on Out-of-School Time). 2008. *Making the case: A 2008 fact sheet on children and youth in out-of-school time.* National Institute on Out-of-School Time at

Wellesley Centers for Women, Wellesley College. http://www.niost.org/pdf/Final2008FactSheet.pdf

Noar, S.M. 2003. The role of structural equation modeling in scale development. *Structural Equation Modeling* 10, no. 4: 622–47.

Patton, M.Q. 1996. *Utilization-focused evaluation: The new century text.* 3rd ed. Thousand Oaks, CA: Sage.

Powell, R., M. Stern, and N. Ardoin. 2006. A Sustainable Evaluation Program framework and its application. *Applied Environmental Education and Communication* 5: 231–41.

Presser, S., M.P. Couper, J. Lessler, E. Martin, J. Martin, J.M. Rothgeb, and E. Singer. 2004. Methods for testing and evaluating survey questions. *Public Opinion Quarterly* 68, no. 1: 109–30.

Rossi, P.H., and H.E. Freeman. 1993. *Evaluation: A systematic approach.* 5th ed. Newbury Park, CA: Sage.

Ruff, H.A., and K.R. Lawson. 1990. Development of sustained, focused attention in young children during free play. *Development Psychology* 26: 85–93.

Satorra, A., and P.M. Bentler. 1994. Corrections to test statistics and standard errors on covariance structure analysis. In *Latent variables analysis*, ed. A. von Eye and C.C. Clogg, 399–419. Thousand Oaks, CA: Sage.

Schafer, J.L., and J.W. Graham. 2002. Missing data: Our view of state of the art. *Psychological Methods* 7: 147–77.

Simmons, B., M. Archie, T. Bedell, J. Braus, G. Holmes, M. Paden, R. Raze, et al. 2004. *Environmental education materials: Guidelines for excellence.* Washington, DC: North American Association for Environmental Education.

Smith, E. P. 2007. The role of afterschool settings in positive youth development. *Journal of Adolescent Health* 41: 219–20.

Stern, M.J., R.B. Powell, and N.M. Ardoin. 2008. What difference does it make? Assessing student outcomes of participation in a residential environmental education program. *Journal of Environmental Education* 39: 31–43.

Tabachnick, B.G., and L.S. Fidell. 2007. *Using multivariate statistics.* 5th ed. Needham Heights, MA: Allyn & Bacon.

Vandenberg, R.J., and C.E. Lance. 2000. A review of and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods* 3: 4–69.

Weiss, C.H. 1998. *Evaluation: Methods for studying programs and policies.* 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Whiteside-Mansell, L., and R.F. Corwyn. 2003. Mean and covariance structure analyses: An examination of the Rosenberg self-esteem scale among adolescents and adults. *Educational and Psychological Measurement* 63: 163–73.

Yuan, K.H., and P.M. Bentler. 2001. Effects of outliers on estimators and test in covariance structure analysis. *British Journal of Mathematical and Statistical Psychology* 54: 161–75.